MARIA BEATRICE MAGRO

Biorobotics, robotics and criminal law

1. Behavioural Genetics, neurosciences and criminal law

The subject of behavioural genetics and neurosciences, within the wider issue of the relations between law and technology, in a legal perspective, poses on the background the question of the conception of corporeity and anthropology underlying legal conceptions. On a penal level, many are the reflections of the recent studies on behavioural genetics, molecular biology and neurosciences, studies which allow us to better understand the decisional and deliberative human process, individuating a biological-chemical basis imposing a re-definition of the notion of free will. The impact of studies of philosophy of mind, biology and neurology reflects on the important categories of criminal law and procedure; it influences categories such as imputability (where the ability to understand includes: empathy, moral thought and counterfactual reasoning, and the ability of willing: the ability of controlling the impulse of movement), the ability to stand in trial, the evaluation of declarative evidence; it is relevant for crimes concerning the abuse of drugs, with preventive purposes, in the executive phase of punishment, whereas the preventive conception of punishment expresses that inhibitory function, of cultural orientation and of veto that influences control functions.

2. Biorobotics, interfaces brain-machine and human strengthening: precautionary philosophy and heuristics of risk contrasting.

The field of robotics and biorobotics is, *par excellence*, interdisciplinary. The most topical aspect of robotics concerns biorobotics, that is the combination or hybridisation man-machine, through the implant of robotic parts in the human body (inoxidizable computers, computer implants in the body, artificial limbs) with therapeutic or physical strengthening purposes. The surprising data, distinguishing man-machine hybridisation from other non-cybernetic prostheses, is the interaction between nervous system, cerebral impulses and prosthesis animation, to the end of give the subject the perception of movement. Biorobotics takes advantage of the so-called «brain-machine interfaces»: channels that offer the chance to influence the mind/conscience states of a person and that allow the transmission of external electrical signals directly to the brain. This way, the bionic limb is capable to recognise the subject's will and to execute the motion orders of brain in real time. The interfaces between human brain and machine consent the reading and use of neural signals associated to cognitive activity in order to control an artificial limb or the trajectory of a mobile robotic platform.

Important ethical questions arise: on a bioethical and philosophical level, the technological evolution questions the concept of "human" and "humanity", of personal identity and even, on a legal level, the concept of responsibility. The real presence of hybridising man-machine phenomena (cyborg) shall make us reflect also on the legal problems connected to the functioning or the disfunctioning of such new technologies and to the legal regime of responsibility for harm or offences committed by hybridized subjects against close relatives and third parties. The criminal liability aspect, in the case of crime commission, raises the following questions: is it possible to establish whether the incident have been caused by a problem of control of the artificial limbs? The functioning of such limbs follows scientific laws? Who is to be considered responsible for the death of a third, when it is not possible to ascertain a construction or functioning fault of the bionic limb? The "action" performed with an artificial limb may be considered "conscious and intentional"?

More in the background are the usual aspects concerning the limits to the possibility of dispose of one's own body and physical integrity, where the subject is intertwined with the distinction between therapeutic intervention (and the respective epistemological statute: free, conscious and informed consent) and improving or strengthening intervention not for therapeutic purposes. The problem is that it is very difficult to trace a conceptual distinction between therapy and strengthening and we may even maintain that it lacks practical meaning. The very same notion of health comes to play: if intended in a subjective sense, as recognised by the World Health Organisation, that is, as a synonym of complete (and utopian) state of well-being, covering not only a physical dimension, but also the psychic dimension, which goes beyond what is necessary to re-establish a state of alteration, such a notion seems to eliminate the conceptual differentiation between therapy and pure strengthening (non-therapeutic).

The question of the moral and legal difference between interventions for therapeutic purposes and interventions for strengthening purposes interferes with the underlying problem: what are the effects of the strengthening technologies? Are they positive or harmful? Such (positive or harmful) effects are limited to the individual or are extended to environment and future generations? The issue of strengthening opens a new frontier of the bioethical debate, raising several ethical worries due to the situation of double epistemic uncertainty that connotes it. The first level of epistemic uncertainty is of a scientific kind. Indeed, on the one hand, they represent relatively new techniques – some merely prospected for the future, some in the phase of first experimentation on their possible uses as for their effectiveness, security, both in the short and in the long period. The second level of uncertainty concerns the evaluation on a moral and ethical level on their possible effects on the individual, future generations and environment. The problem is how to evaluate, from an ethical perspective, the refinement of techniques of strengthening in terms of "authenticity" (is it more authentic the life of an individual who improved his own potential or that of an individual living in the forests?), of dignity (but what concept of human dignity is to be assumed, as an objective of realisation or as naturality?).

In the current debate, those articulating theories and arguments in defence of strengthening (libertarian theory, utilitarian theory, so-called "technophiles" Nicholas Agar, Allen Buchanan, Nick Bostrom, John Harris, Julian Savulescu) are opposed to the detractors or so-called "bioconservative" (Francis Fukuyama, Jurgen Habermas, Leon Kass, Michael Sandel) who analyse the possible threats to man and future generation. In such a perspective, the problem of strengthening is dealt with in the field of the theories of justice, which specific reference to the problem of inequality (strengthened/non strengthened), of the possible reflections on the access to strengthening, on the non-acceptation of disability (given the ever-growing spread between the dis-able, the able, the super-able and the strengthened).

Thus, the problem is to detect a model of legal regulations in situations of scientific uncertainty. The hypothesis of the ultra-prudential prohibition of behaviours having positive effects comes to play. If the potential effects of bionic and biorobotic strengthening technologies are harmful (for the individual, for future generations or for the environment) or partially positive (for the individual) but harmful for the future generation and for the environment, the possible criminalisation of biorobotic technologies of human strengthening finds its foundations in the classic paradigm of legitimacy of the prohibition of causing harm to others or to self. Nevertheless, if, as we have been saying, there is a total scientific uncertainty on the harmfulness of such interventions, and only their positive effect under the psycho-physical aspect is ascertained, even if with mere regard to the single individual upon which they are practised, then the problem is whether, from an ethical perspective, the bio-strengthening should not represent a moral and social duty (especially if their effects are positive for everybody, both on a personal and general level). The hypothetical and future criminalisation of such technologies would in that case reflect an inversion of the strong and weak paternalistic paradigm: it is forbidden what is good for man on the premise that strengthening is an expression of futile motives. This goes beyond the prohibition of causing harm to self or to others: one can almost see a norm of ultra-precautionary prohibition which goes beyond the model

of justification of strong paternalism and which incorporates the heuristic of strong precaution.

3. Robotics, drones and Artificial Intelligences

Robotics includes the study of Artificial Intelligences, that is, the construction of machines capable to feel, to adapt to environment, to learn, to evolve and even "capable of empathy". Such machines have the following characters: they are interactive and reactive to environment, they act autonomously and unpredictably and in a non-determined way, flexible, easily influenced; they have a system of automatic learning and are therefore given with autonomy, interactivity, adaptability and they are capable to improve their performances. Complex organisations recur to intelligent computer and technologic systems capable to take autonomous decisions (decisions which do not require any explicit authorisation by a human being) and operating as a support for managers especially in the management of technological infrastructures at high risk. Artificial organisations are structures modelled on human ones, in which each intelligent agent, named "personoid" occupies a precise role (including information access, duties, responsibility) and produces different decisional processes.

The fields of application of Artificial Intelligences are the most disparate: there are artificial agents and artificial neural nets in the economic field (robots act as intelligent economic agents without emotions, thus realising the conditions for competitive balance in demand and offer, or the model of perfect distributive efficacy), creation of service robots and assistance robots (robot carers). In medicine, with the development of mini-invasive surgery, robots are used in the so-called telesurgery. Wide applications are made within the military field, with the use of drones as weapons or robot soldiers. In this regard, many scientists call for prohibition or caution in using autonomous lethal robots, as intelligent military systems making autonomous decisions, the consequences of which may be devastating. Such robot weapons may trigger lethal weapons without human intervention in the decisional process. The International Criminal Court has declared the unlawfulness of the use of robot weapons and drones.

Robot behaviour may be unpredictable, insofar as Artificial Intelligences have an ability of automatic learning, of adapting to environment, of decision and creativity amongst a range of choices. The ways of automatic learning do not always consent to verify whether the robot has really learned (even with a satisfactory approximation) what we want to teach it. There are, therefore, problems of evaluation of the results obtained from the processes of automatic learning by robots. Moreover, the interaction between such an open conditioned system of learning typical of Artificial intelligences and its environment (a-synchronic communication) is often subject to temporal bonds that do not consent to ensure the presence of a human being in the control cycles. Therefore, they are not entirely controllable by man.

The studies of formal methods of verification of the software arise from the need to verify if any execution of a certain program for a calculator satisfies some fundamental requirements. More recently, such methodologies have been extended to the problem of specify and verify the proprieties of systems, generally called "hybrids", which include many typologies of robotic systems. Some limitative results obtained with regard to hybrid systems shows that, in general, it is not possible to verify with such methodologies whether a robotic system satisfy or not certain spatial or temporal bonds in the execution of a particular duty.

What are the practical normative implications of the epistemological reflections on our limited ability to predict the machines' behaviour? Is the producer or the programmer of an intelligent system that learns from experience is able to precisely predict its behaviour, also in the normal conditions of use, and is he therefore responsible? There are two kinds of problems. The first is a philosophical one, it is the question of the ontological statute of particularly evolved machines. Are they simple objects or do they cross the threshold of minimum requisites to be recognised with some level of subjectivity? The second is practical-functional-legal, and it is the question whether it is convenient and useful to legal purposes attribute to such artefacts some level of subjectivity. The

problem consists therefore in establishing whether modern technology created a new typology of subject, non-human subject, non-human agent. Such Artificial Intelligences, as provided with autonomy, unpredictable agents are really capable to manifest their "own" intentionality and therefore their own acting? Are they mere instruments, or are they subjects, also mediated respect to another author?

According to the orientation called Strong Artificial Intelligence calculators are machines able to express authentic thought and to produce intellectual processes identical to the human ones. There would not be any ontological-qualitative difference between human brain and electronic brain and between human intelligence and artificial intelligence, the only difference would be the seat or the physical support, the human head made of meat, bones and other biological materials and the structure of a calculator made of metal and energy. The structure does not count, the function does. There are intelligences themselves, independently from the physical seat in which they reside (functionalist thesis). The contrary thesis, called Weak Artificial Intelligence, or thesis of the non-authenticity of mechanical thought and of the ontological diversity between artificial intelligence and natural intelligence, maintains that machines emulate and replicate the human intellectual processes only of which they represent copies; what does matter is the structure and not the function, that is the seat where the activity takes place and not how it develops. Two opposites conceptions of the ontological statute of machines derive from these two visions.

A) Criminal liability aspects: the fantastic hypothesis of the robot as a non-human agent penally responsible, new subject of criminal law. Such a hypothesis raises problems of definition of subjective capacity, of the notion of action, of the notion of culpability: are robots blamable and motivable? They express their own will, may they be "authors" of crimes? The themes presents an analogy with that of the penal-administrative responsibility of collective bodies and conflicts with an anthropomorphic conception of criminal law; collective entities have no body nor soul, but they are subjected to criminal law; robots have a "body", a matter upon which the penal sanction could fall (for instance the deactivation of the machine or its destruction), but they have not a soul, albeit if provided with autonomous decisional capacity and being able to interact with environment. But is the acting of a robot a "real" penally relevant acting, could it be defined conscious and voluntary?

B) Hypothesis of the criminal responsibility of the physical person: if robots are means and not non-human agents. Believing that robots are not agents in the strict sense poses the problem of the exclusively human responsibility of the programmer or user. To which title may we attribute any responsibility to man, if we assume that the intelligent computer system is a decision maker? Intelligent robots act in a non-programmed and unpredictable way. Such unpredictability poses problems of attribution of penal responsibility against programmers, producers and users. An inverse problem of imputation human agent-collective entity: the human agent acts, the entity responds. Here the non-human agent acts, while the physical person responds.

The criminal liability of the programmer or user: which model of penal responsibility against the human agent? Direct or indirect responsibility, intentional or negligent? There are two solutions: A) intentional, because the "voluntary" action of the non-human agent determine the man's responsibility and because the action of the robot identifies with and represent a sort of *longa manus* of the human person (theory of organic identification: the robot). B) Negligent, individuated in the fault of programming, construction, choice, use, maintenance, functioning of intelligent robots.

The negligence of the programmer: it is necessary to elaborate a new normative concept of culpability that consents to elaborate another concept of "negligence for programming", whereas one can prospect a cause for exclusion of culpability consisting in the predisposition of security measures to prevent the realisation of crimes by intelligent robots (which may practically enact the three laws of robotics by Asimov).

Furthermore, it will be necessary to individuate the precautionary rules concerning the programming, the choice, the use, the control and the maintenance of robots. Limits to the responsibility of programmers, producers and users shall be individuated in case of improper use of

robot or in case of force majeur or unforeseen accident.

When the artificial agent is evolved and benefits of a margin of creativity and decisionality (it is a decision maker) and there is no visible fault in control, use, maintenance of the robots, the subjective imputation of the crime to the physical person will be intentional or negligent? Anglo-Saxon literature talks of foreseeability and logic foreseeable development, similarly to art. 116 of Italian penal code concerning the responsibility of the participant to the crime for a different, more serious event than the wanted event, as long as foreseeable, and even if not all the rings of the causal chain of production of the event. The difference is that the programmer or the user does not concur in the realisation of a base-crime and therefore the rule of *versari in re illicita* does not apply, but create a condition of lawful tolerated risk. The foreseeability as component of guilt, based on scientific laws, even if not corroborated by studies and consolidated, or in the total ignorance of scientific laws of causal explanation. In the very end, the behaviour of the programmer who does not predict the unpredictability of the intelligent robot is negligent! Even abstract and generic foreseeability of future damages not well identified yet triggers negligent responsibility, even when not all the rings of the causal process are known: risk of identifying negligence with precaution and prudentialism.